International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Behavioral Twin AI: A New Paradigm for Return Fraud Prevention

Author

Sandnya Dalvie

Abstract

Retail return fraud has evolved into a sophisticated, high-impact global threat, costing retailers tens of billions of dollars annually worldwide. As fraudsters exploit lenient return policies, cross-border commerce, and digital loopholes, traditional rule-based fraud detection systems increasingly struggle to differentiate malicious intent from legitimate customer behavior.

Behavioral Twin AI introduces a paradigm shift in fraud prevention by modeling each customer's unique behavioral patterns over time. Rather than applying generic thresholds or static rules, the system continuously learns what is normal for each individual and identifies risk through meaningful deviations across behavior, device, location, channel, and intent.

This whitepaper presents the Behavioral Twin AI framework, demonstrates its application through a real-world scenario, and quantifies its business impact. By combining behavioral intelligence, computer vision, and natural language processing, organizations can materially reduce fraud losses while preserving customer trust, operational efficiency, and customer experience on a global scale.

The Billion Dollar Problem

A 2025 global retail fraud survey found that over 60% of retailers worldwide now cite return fraud as their largest preventable source of loss, contributing to tens of billions of dollars in annual revenue leakage globally. Fraud techniques such as wardrobing, refund abuse, receipt manipulation, and device hopping have become increasingly sophisticated and are deliberately engineered to evade static, rule-based detection systems.

At the same time, overly aggressive fraud controls disproportionately impact genuine customers, resulting in blocked returns, delayed refunds, manual reviews, and erosion of brand trust. As customer experience becomes a competitive differentiator, retailers face a critical challenge: how to stop fraud decisively without alienating loyal, high-value shoppers.

Limitations of Traditional Fraud Detection

Legacy fraud prevention relies heavily on deterministic logic: if-then rules, blacklists, and coarse scoring. These systems lack behavioral context and fail to understand intent. Fraudsters adapt faster than static rules can be updated, while genuine customers are penalized for occasional anomalies.

Key limitations include:
- No individualized customer baselines
- Inability to contextualize anomalies
- High false-positive rates
- Limited explainability for investigators

Proposed Solution: Introducing Behavioral Twin AI

Behavioral Twin AI creates a living, continuously evolving digital representation of each customer derived from their historical interactions across the retail ecosystem. This behavioral model captures patterns spanning purchase

frequency, spend velocity, return behavior, device usage, geographic consistency, channel preferences, and interaction timing. Unlike traditional fraud systems that rely on static profiles or population-level averages, the Behavioral Twin establishes a personalized baseline that reflects how each individual customer normally behaves over time.

Instead of comparing customers to generic thresholds or peer cohorts, every new transaction or return request is evaluated against that customer's own behavioral baseline. Actions are assessed in context across multiple dimensions, including temporal patterns, device and location continuity, order composition, and stated intent. This allows the system to distinguish between legitimate behavioral variation and meaningful anomalies that indicate elevated fraud risk, even when individual signals appear benign in isolation.

This approach mirrors how humans intuitively identify suspicious behavior — not by judging single actions independently, but by recognizing deviations from established patterns of conduct. A high-value purchase, a new device, or a sudden location change may be acceptable individually; however, when such signals occur simultaneously and conflict with a customer's historical behavior, they represent a materially higher risk. Behavioral Twin AI captures and quantifies these deviations with machine precision and consistency.

By modeling behavior dynamically rather than enforcing rigid rules, Behavioral Twin AI adapts as customers' habits naturally evolve. Life events, seasonal changes, and channel shifts are learned over time, reducing false positives while maintaining sensitivity to genuine abuse. The result is a fraud prevention system that is adaptive, explainable, and resilient against increasingly sophisticated return fraud tactics — protecting revenue without compromising customer experience.

How Behavioral Twin AI Works

Behavioral Twin AI operates as a multi-layered intelligence pipeline designed to evaluate customer behavior holistically and in real time. Rather than relying on a single signal or static rule, the system integrates multiple data sources, analytical layers, and machine learning models to produce a contextual and explainable risk assessment for every transaction or return event.

1. Data Collection

The pipeline begins with comprehensive data ingestion across the customer journey. This includes transactional history (purchases, returns, refunds), geolocation data, device and browser metadata, account changes, stated return reasons, and supporting artifacts such as product images. By capturing both behavioral and contextual signals, the system ensures that decisions are grounded in a complete view of customer activity rather than isolated events.

2. Feature Engineering

Raw data is transformed into behaviorally meaningful features that describe how a customer typically acts. Examples include spend velocity, order value variance, device and location consistency, time-to-return patterns, return frequency by category, and channel switching behavior. These features establish the customer's behavioral baseline and allow the system to measure the magnitude and significance of deviations from normal behavior.

3. Machine Learning Layer

Three specialized machine learning models operate in parallel. The Behavioral Model analyzes deviations from historical patterns and peer-agnostic baselines. The Visual Model applies computer vision techniques to assess product images for authenticity, condition, and SKU consistency. The Textual Model uses natural language processing to evaluate return descriptions for generic, abusive, or high-risk language patterns commonly associated with fraud.

4. Risk Scoring and Decisioning

Outputs from each model are normalized and combined into a unified risk score in real time. This score reflects both the severity of behavioral deviation and the confidence of supporting evidence. Based on configurable thresholds, the system can approve returns seamlessly, route cases for manual review, request additional verification, or temporarily pause high-risk refunds — all with clear, auditable reasoning.

5. Continuous Learning and Feedback

Behavioral Twin AI continuously improves through feedback loops. Confirmed fraud cases, investigation outcomes, customer appeals, and false positives are fed back into the training process. This enables the system to adapt to emerging fraud tactics, seasonal behavior changes, and evolving customer norms, ensuring long-term accuracy, resilience, and reduced operational friction.

Working Example: Sarah's Case

Sarah is a long-standing, high-trust customer with a three-year purchase history and a well-established behavioral profile. She consistently spends an average of £800 per month across apparel and home goods, with purchasing patterns that show stable order values, predictable timing, and limited variance. Her return rate of approximately 8% aligns with typical retail behavior and is largely driven by legitimate factors such as sizing adjustments or color preferences rather than abuse. Over time, her Behavioral Twin has learned a reliable baseline characterized by consistent device usage, stable geographic location, and a regular purchase cadence across web and mobile channels.

On one Tuesday morning, however, Sarah's account begins to exhibit a sequence of behaviors that materially deviate from her historical norms. Within a short time window, four high-value apparel purchases are placed, totaling £3,200—representing a fourfold increase over her typical monthly spend. While high-value purchases alone are not inherently suspicious, the sudden spike in spend velocity is inconsistent with Sarah's established buying behavior.

Compounding the anomaly, the orders are placed from a previously unseen device and from a geographic location that differs from her historical shopping patterns. Shortly before the purchases, the shipping address on the account is changed, introducing additional risk related to account access and fulfillment. Within 24 hours of delivery, a return is initiated via the mobile application, significantly faster than Sarah's typical return timing and inconsistent with her historical decision-making patterns.

Individually, each of these signals could be explained by benign circumstances. However, when evaluated collectively against Sarah's Behavioral Twin, the convergence of elevated spend, new device usage, location change, address modification, and accelerated return timing represents a high-confidence behavioral deviation. This contextual assessment allows the system to flag the activity as elevated risk while preserving explainability and avoiding blanket assumptions about customer intent.

7. Risk Evaluation and Outcome

The Behavioral Twin AI system evaluates the event in real time, completing its full analysis within milliseconds of the return request being initiated. Rather than relying on a single indicator, the system orchestrates multiple intelligence layers that independently assess risk and collectively inform the final decision. This layered evaluation ensures both accuracy and resilience against isolated false signals.

The **Behavioral Analysis layer** compares the current activity against Sarah's established Behavioral Twin. It identifies a statistically significant deviation across multiple dimensions, including spend velocity, order value distribution, device continuity, geographic consistency, shipping address stability, and return timing. While none of these signals alone would necessarily indicate abuse, their simultaneous occurrence and magnitude relative to Sarah's historical norms materially elevate risk confidence. This layer contributes the largest weighting to the overall score, as it directly measures deviation from trusted behavior.

The **Image Verification layer** analyzes the product images submitted with the return request using computer vision techniques. The system evaluates image authenticity, checks for inconsistencies between the returned item and the original SKU, and assesses indicators of prior use or substitution. Patterns commonly associated with wardrobing and item swapping—such as altered packaging, wear indicators, or mismatched identifiers—are flagged and incorporated into the risk assessment with quantified confidence scores.

In parallel, the **Natural Language Processing (NLP) layer** evaluates the textual return reason provided by the customer. The model detects generic, low-effort, or templated language that frequently appears in abusive return patterns, as well as inconsistencies between stated reasons and historical behavior. While NLP signals are weighted conservatively to avoid penalizing legitimate customers, they provide valuable corroborating evidence when combined with behavioral and visual findings.

The outputs from all three layers are normalized and aggregated into a unified risk score of **86 out of 100**, exceeding the high-risk threshold. Based on this classification, the return is temporarily paused and routed for manual review rather than automatically rejected. Investigators are presented with clear, explainable signals detailing which behavioral deviations and supporting evidence contributed to the decision, enabling faster resolution, consistent judgment, and reduced customer friction.

Explainability and Trust

Unlike traditional black box scoring systems that produce opaque risk outcomes with little context, Behavioral Twin AI is designed with explainability as a core architectural principle. For every decision, the system generates transparent, human-readable explanations that clearly articulate which behavioral signals deviated from an individual customer's established baseline and how those deviations contributed to the overall risk assessment. Investigators can see, in precise and auditable terms, factors such as abnormal spend velocity, new device usage, geographic inconsistencies, accelerated return timing, or supporting visual and textual indicators, along with their relative impact on the final score. This level of clarity significantly reduces investigation time, improves decision consistency across teams, and minimizes unnecessary escalation of false positives. Equally important, explainability builds internal confidence in automated decisioning, supports regulatory and compliance requirements, and enables customer-facing teams to resolve disputes with factual, defensible reasoning, strengthening trust between the retailer, investigators, and legitimate customers.

Industry Applications

Behavioral Twin AI is applicable across multiple industries:

- Retail & E-commerce: Return fraud, promotion abuse, account takeover
- Financial Services: Transaction anomaly detection, synthetic identity fraud
- Travel & Hospitality: Booking abuse, chargeback prevention
- Marketplaces: Seller-buyer trust scoring and dispute resolution
- Subscription Services: Free-trial abuse and churn risk management

The Behavioral Twin Advantage

Behavioral Twin AI delivers three core advantages that fundamentally differentiate it from traditional, rule-based fraud prevention systems and enable organizations to combat return fraud without compromising customer experience.

Real-Time Adaptive Response allows the system to detect and act on fraud risk before financial loss occurs. By evaluating behavioral signals at the moment a return or refund is initiated, and continuously learning from confirmed outcomes, the platform adapts to emerging fraud tactics, seasonal behavior shifts, and evolving customer patterns—preventing abuse before refunds are processed rather than reacting after losses are incurred.

Personalization ensures that every customer is assessed against their own historical behavior instead of generic thresholds or population-wide averages. This individualized baseline dramatically reduces false positives by recognizing legitimate behavioral variation while remaining highly sensitive to meaningful deviations that indicate risk. High-value, loyal customers are protected from unnecessary friction, while sophisticated fraud attempts that mimic "normal" behavior at an aggregate level are more effectively detected through personalized context.

Explainable AI provides transparent, auditable reasoning behind every risk decision. Investigators are presented with clear explanations detailing which behavioral, visual, or textual signals contributed to risk elevation and how they

were weighted. This improves investigator confidence, accelerates case resolution, supports regulatory and compliance requirements, and enables customer-facing teams to communicate decisions with clarity and fairness—ultimately strengthening trust across internal teams and with legitimate customers.

Conclusion

As return fraud continues to grow in scale and sophistication, retailers can no longer rely on static detection methods built on fixed rules, thresholds, and historical assumptions. Modern fraud schemes are adaptive by design, intentionally mimicking legitimate customer behavior to evade traditional controls, while rapidly evolving in response to policy changes and enforcement tactics. In this environment, rule-based systems become increasingly brittle—either lagging emerging threats or overcorrecting in ways that introduce friction for genuine customers and erode trust.

Behavioral Twin AI represents a fundamental shift in how fraud is understood and managed, moving the focus from rule enforcement to behavioral intelligence. By modeling each customer's unique behavioral patterns over time and evaluating actions in context, the system detects risk through meaningful deviations rather than isolated signals. This approach enables retailers to identify sophisticated abuse that would otherwise appear normal under generic rules, while simultaneously accommodating legitimate changes in customer behavior as life circumstances, channels, and preferences evolve.

Crucially, this behavioral understanding allows fraud prevention to operate in harmony with customer experience rather than in opposition to it. High-trust customers are protected from unnecessary friction, refunds are processed efficiently when risk is low, and investigative resources are focused on where they add the most value. By preventing losses before they occur and preserving seamless interactions for legitimate shoppers, Behavioral Twin AI enables retailers to safeguard revenue, reduce operational cost, and strengthen long-term customer loyalty in an increasingly competitive retail landscape.

References:

https://www.businessinsider.com/return-fraud-amazon-shipping-retail-theft-wardrobing-online-shopping-2025-7?utm_source

https://arxiv.org/abs/2505.10050?utm_source

https://www.sciencedirect.com/science/article/pii/S2665917424001144?utm_source#sec3

https://ieeexplore.ieee.org/document/10506811